

INICIAR SESIÓN

NUESTROS PLANES

TODOS LOS
CURSOS

FORMACIONES

CURSOS

PARA
EMPRESAS

ARTÍCULOS DE TECNOLOGÍA > DATA SCIENCE

PLN: ¿Qué es el procesamiento del lenguaje natural?



goncalves-thiago

09/01/2021

En este artículo descubrirás **qué es PLN, orígenes y desafíos**.

El [Procesamiento natural del lenguaje](#) (*PLN* o *NLP*) mezcla **Ciencia de la computación**, **inteligencia artificial** y **lingüística** dedicándose a la generación y comprensión automática del lenguaje natural. Pero ¿qué pasa con el día a día dónde encontramos PLN? Fuente: Aliz.ia

¡PLN en el día a día!

¿Cuántos mensajes de texto has leído hoy? ¿Cuántos audios se enviaron por las aplicaciones de mensajería? ¿Cuántos correos electrónicos se recibieron? ¿Cuántas publicaciones en redes sociales se enviaron?

Para todas las preguntas anteriores, la respuesta probable es que hubo muchas. [webfx](#) muestra en tiempo real cuántos datos se están traficando en internet en un momento dado, si ingresas notarás que es un volumen gigantesco, gran parte de estos datos son texto, audios o videos.

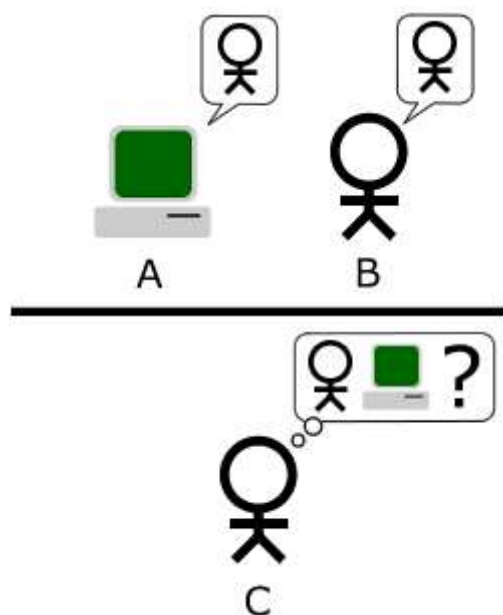
La disponibilidad de datos en lenguaje natural nunca ha sido tan grande, por lo que tratar y desarrollar algoritmos que permitan el desarrollo de aplicaciones útiles en nuestro día a día con toda esta información es parte fundamental del procesamiento del lenguaje natural.

Pero queda una pregunta, ¿de dónde vino esta idea de mezclar la computación con lingüística?

Origen

En 1950 **Alan Turing** publicó el artículo [Computing Machinery and Intelligence](#), siendo pionero en inteligencia artificial, trayendo la propuesta de una prueba actualmente conocida como [Prueba de Turing](#).

El desafío de identificar una máquina o una persona basándose únicamente en las preguntas y respuestas de una conversación entre hombre-máquina, dio lugar al procesamiento de lenguaje natural tal como lo conocemos hoy.



Fuente: Wikipedia

Luego, en la década de 1960, surgieron los primeros sistemas exitosos, entre los que destaca el [ELIZA](#) que simula un psicoterapeuta.

Usando información sobre el pensamiento y las emociones humanas, ELIZA creaba interacciones sorprendentemente humanas. Cuando el usuario extrapolaba la base de conocimientos del programa, ELIZA proporcionaba respuestas genéricas para mantener la conversación aún más humana.

Hasta la década de 1980, la mayoría de los sistemas PLN se basaban en complejos conjuntos de reglas y, a fines de la década, [machine learning](#) iniciaron una nueva y promisoría fase en el área, donde los algoritmos eran relativamente simples, como un árbol

de decisiones y ya producían sistemas de reglas tan similares o incluso mejores que las reglas escritas “a mano” a principios de la década.

Evolución de modelos

Desde entonces, modelos estadísticos, técnicas de machine learning y más recientemente **deep learning** dominan el área, creando sistemas cada vez más cercanos a pasar la prueba de Turing.

Los nuevos modelos nos impresionan con traductores cada vez mejores, **chatbots** humanizados y clasificadores capaces de percibir sentimientos sobre un producto analizando sus valoraciones textuales.

Un caso muy interesante ocurrió en septiembre de 2019, cuando el Instituto Allen de Inteligencia Artificial (AI2) reveló al mundo el programa llamado [Aristo](#) que puede resolver correctamente una prueba científica con una tasa de acierto superior al 90%, aunque la prueba es para noveno año en Brasil y parece relativamente simple para un adulto, esta tarea es extremadamente difícil para una computadora.

Para encontrar las respuestas correctas a cada una de las preguntas, fue necesario analizar miles de millones de documentos. Extraer información útil de estos miles de millones de documentos solo fue posible con la intersección de áreas como la computación, la lingüística y la inteligencia computacional.

PLN es justamente esa área de conocimiento interdisciplinar que trabaja con la interacción entre computadoras y humanos utilizando el lenguaje natural. A partir de modelos lingüísticos como la teoría gramatical generativa de [Noam Chomsky](#) y modelos estadísticos, PLN procesa, analiza y genera valor utilizando datos en lenguaje natural, ya sean textos o audios, como entrada.

¿Cuáles son los retos?

El procesamiento del lenguaje natural no es un problema resuelto, especialmente cuando se trata de **análisis semántico**. En el contexto de la lingüística, la semántica es el estudio del significado que utiliza el ser humano para expresarse a través del lenguaje. Para entender la ironía o el sarcasmo, por ejemplo, no solo se analizan las palabras utilizadas en la frase, sino el contexto en el que se aplican.

La generalización de **Modelos de NLP** también es un problema, aunque está dando grandes pasos debido a las técnicas de Deep Learning más avanzadas lanzadas en los últimos años, crear modelos entrenados en un lenguaje que sean genéricos o suficientes para ser aplicados satisfactoriamente en otros idiomas sigue siendo un problema, desafiante y importante, ya que no todas las lenguas tienen una base de datos lo suficientemente buena para entrenar los modelos de Deep learning más avanzados.

La especificidad de cada lengua hace que trabajar con el lenguaje natural sea desafiante y aún lleno de retos que, una vez superados, allanarán el camino para el desarrollo de muchas aplicaciones con el potencial de transformar la forma en que manejamos la tecnología.

Conclusión

El procesamiento del lenguaje natural es esta área que mezcla la computación, la inteligencia artificial y la lingüística, llena de desafíos, con aplicaciones que son parte de nuestra vida diaria y un mundo de posibilidades con el potencial de transformar la forma en que usamos teléfonos celulares, computadoras y otras tecnologías. ¿Te gustó este artículo y quieres saber más sobre PLN? Aquí en **Alura** tenemos mucho contenido de [Data Science](#) para ti.

Puedes leer también:

- [Buscando tweets con Python](#)
- [¿Cómo comparar objetos en Python?](#)
- [Matplotlib una biblioteca Python para generar gráficos interesantes](#)

ARTÍCULOS DE TECNOLOGÍA > DATA SCIENCE

**En Alura encontrarás variados cursos sobre Data Science.
¡Comienza ahora!**

SEMESTRAL

US\$49,90

un solo pago de US\$49,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 6 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

ANUAL

US\$79,90

un solo pago de US\$79,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 12 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

Acceso a todos
los cursos

Estudia las 24 horas,
dónde y cuándo quieras

Nuevos cursos
cada semana

NAVEGACIÓN

PLANES

INSTRUCTORES

BLOG

POLÍTICA DE PRIVACIDAD

TÉRMINOS DE USO

SOBRE NOSOTROS

PREGUNTAS FRECUENTES

¡CONTÁCTANOS!

¡QUIERO ENTRAR EN CONTACTO!

BLOG

PROGRAMACIÓN

FRONT END

DATA SCIENCE

INNOVACIÓN Y GESTIÓN

DEVOPS

AOVS Sistemas de Informática S.A
CNPJ 05.555.382/0001-33

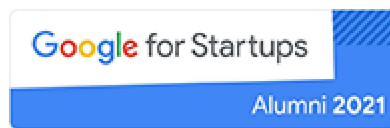
SÍGUENOS EN NUESTRAS REDES SOCIALES



ALIADOS



En Alura somos unas de las Scale-Ups seleccionadas por Endeavor, programa de aceleración de las empresas que más crecen en el país.



Fuimos unas de las 7 startups seleccionadas por Google For Startups en participar del programa Growth Academy en 2021

POWERED BY

CURSOS

Cursos de Programación

Lógica de Programación | Java

Cursos de Front End

HTML y CSS | JavaScript | React

Cursos de Data Science

Data Science | Machine Learning | Excel | Base de Datos | Data Visualization | Estadística

Cursos de DevOps

Docker | Linux

Cursos de Innovación y Gestión

Productividad y Calidad de Vida | Transformación Ágil | Marketing Analytics |
Liderazgo y Gestión de Equipos | Startups y Emprendimiento